Unraveling the Signature genes involved in tumors of epithelial and mesenchymal origin

Thesis submitted to
National Institute Of Technology, Rourkela
For The Partial Fulfillment Of The Master Degree In Life Science

Submitted By:.

Bini Chhetri Soren

411LS2067

Under the guidance of:

Dr. Bibekanand Mallick

Department of Life Science

National Institute of Technology Rourkela

Rourkela – 769 008

NATIONAL INSTITUTE OF TECHNOLOGY, ROURKELA
राष्ट्रीय प्रौद्योगिकी संस्थान, राउरकेला

Dr. Bibekanand Mallick, **M.Tech., Ph.D.**
Assistant Professor

RNA Biology & Functional Genomics Lab.
Department of Life Science
National Institute of Technology
(*Ministry of H.R.D, Govt. Of India*)
Rourkela - 769 008, Odisha, India

Telephone: +91-661-246 2685 (O)
E-mails: vivek.iitian@gmail.com, mallickb@nitrkl.ac.in
Homepage: http://vvekslab.in

Date: 10. 05. 2013

CERTIFICATE

This is to certify that the thesis entitled "Unraveling the Signature genes involved in tumors of epithelial and mesenchymal origin" submitted to National Institute of Technology; Rourkela for the partial fulfillment of the Master degree in Life science is a faithful record of bonafide and original research work carried out by Bini Chhetri Soren under my supervision and guidance.

(Dr. Bibekanand Mallick)

ACKNOWLEDGEMNT

I would like to express my profound gratitude to my supervisor Dr. Bibekanand Mallick, Department of Life Science, National Institute of Technology, Rourkela for his constant guidance, support, encouragement and affection.

I am also heartily thankful to Dr. Sameer Kr. Patra (HOD), Dr. Sujit Kr. Bhutia, Dr. Bismita Nayak, Dr. Rasu Jayabalan and Dr. Suman Jha for their inspiring suggestion, valuable advice and cooperation.

I am highly obliged and grateful to Ms Devyani Samantarrai and Ms Debashree Das (Ph.D Scholars) for their support, suggestions and timely advice during the course of my work.

I would also like to thank my lab mates and friends for their help and moral support during the period of my work.

I express my heartiest gratitude to my beloved parents for their encouragement, moral support, love and blessings without which the present study would not have been successful.

Date: 09 May, 2013

Rourkela Bini Chhetri Soren

List of Symbols and Abbreviations Used

µg	Micro gram
µl	Micro litre
°	Degree
C	Centigrade
ml	Mili litre
mM	Mili molar
min	Minutes
%	Percentage
No.	Number
FC	Fold Change
Reg.	Regulation

TABLE OF CONTENT

Sl. No.	CONTENTS	PAGE No.

LIST OF TABLES

LIST OF FIGURES

1. ABSTRACT:

Tumors mainly originate from epithelial and mesenchymal cells, majority of which are from epithelial origin. There are certain mechanism like EMT and MET which plays a critical role in malignant tumors or cancer progression. There are various other phenomenon and mechanisms occurring at the molecular and cellular level which are yet to be discovered. Success in unraveling such a mystery can give clue to understand the fatal disease like cancer. Roles of genes involved in various cancers and their expression in different forms of cancer as well as in normal condition can give an insight about cancer. This study could be carried out using microarray analysis. Present work mainly focuses to understand the similarity and dissimilarity between the tumors of epithelial and mesenchymal origin by microarray analysis of differentially expressed genes through Genespring software. Gene ontology analysis was also carried out through Genomatix software and two genes (ABCA8 and SMC4) were selected for experimental validation based upon their regulation and association with cancer. Expressions of these two genes were experimentally validated through qRT-PCR with respect to housekeeping gene beta-actin in the HeLA cell line, a cervical cancer cell line. Studies like this can help us to understand the behavior of different types of cancer at the molecular and cellular level, thereby developing an effective treatment measure for deadly disease like cancer.

Keywords: Epithelial tumor, mesenchymal tumor, cancer

2. INTRODUCTION:

In this era of globalization with the aid of technological revolution the human society strive to make the overall quality of life better. But still there are some diseases which persist as a mystery yet to be unsolved. One such mystery is the mystery of tumors that may further lead to the deadly disease "CANCER" which is haunting every levels of the society around the world. Understanding and unraveling the complexity behind these tumors is a primary concern among the scientific community in present condition.

Tumors are an abnormal mass of tissue which is solid or fluid-filled. They can either be benign (non cancerous) or malignant (cancerous). Benign tumors are restricted to grow at one place and they don't invade the nearby tissue, whereas malignant tumor invade the adjacent tissues and show properties of metastasis i.e. the abnormal cell move from one area to another within the patient's body. Therefore, malignant tumors shows high degree of aggressive growth compared to benign tumors (Weinberg, 2006).

According to the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3), tumors are classified based upon the morphology and topography of the neoplasm into epithelial, connective tissue, nervous system and hematologic.

Epithelial tumors begin in the skin or in the tissues that line or cover internal organs. They account for majority of tumors occurring all round the globe and mainly affect the organ systems like digestive, reproductive system etc. They are also known as carcinomas when they exhibit cancerous property.

Mesenchymal tumors originate from the connective tissues, hence are also known as connective tissue tumors. Connective tissues include fat, tendons or other tissues that connect, surround or support the organs in the body. They can arise anywhere in the body but are most commonly found in arms, legs, retroperitonium (area behind the abdomen) and chest. When they show cancerous property, they are known as sarcomas. Carcinomas account for 80 to 90 percent of all cancer and sarcomas being relatively rare malignant tumors comprise less than 10% of all cancers (Jemal et al., 2003).

Tumors of nervous system or neuroepithelial origin are derived from neurons, astrocytes, ependyma, oligodendrocytes or primitive embryonal cells. The most common type of tumor of nervous system is glioblastoma which is highly malignant astrocytic glial tumors with a fast pace of growth.

Hematologic tumor includes leukemias, lymphomas and other related disorders. Leukaemias are cancers that affect cells of blood and bone marrow. It starts in the bone marrow where blood cells develop and generally white blood cells undergo cancerous change. These cells multiply in a uncontrolled manner and affects the ability to synthesize normal cells. These abnormal cells are called blast cells or leukaemic blasts which come out of bone marrow and spread in the body through bloodstream.

Lymphomas are a cancer that affects cells of immune system or lymphatic system known as lymphocytes. There are mainly two types of lymphomas Hodgkin and non-hodgkin. Hodgkin lymphoma is marked by the occurrence of Reed-Sternberg cell and restricted to B-cells whereas non-hodgkin lymphoma can occur in either B-cell or T-cell and is further divided into indolent i.e. slow growing and aggressive i.e. fast growing. Both types of lymphomas can occur in adult as well as children.

Out of all these tumors majority of the tumor originate from the tissues of epithelial origin. Then by the mechanism of epithelial to mesenchymal transition (EMT) they spread to other parts of the body leading to carcinoma invasion and metastasis. And when they reach different sites after metastasis these cells undergoes mesenchymal to epithelial transition (MET) mechanism for clonal outgrowth. By following these two mechanism i.e. EMT and MET the initiation as well as completion of the cancer metastasis is carried out (Chaffer and Weinberg, 2011). EMT which is a conserved cellular program was initially recognized during stages of embryonic development. It has recently been found to have a role in promoting carcinoma invasion and metastasis (Yang and Weinberg, 2008). This mechanism mainly allows the conversion of immotile epithelial cells to motile mesenchymal cells. By studying all these mechanisms and phenomenon we can think that how the behavior of one type of tissue influence the other tissue during cancer progression. And further if we think about the processes occurring at the cellular and molecular level the degree of complexity increases. The mechanism of EMT and MET are one such example among the various other processes occurring at the molecular level some of which are yet to be understood and discovered.

Proper understanding and unveiling the mystery of the tumors particularly of these two origins i.e. epithelial and mesenchymal origin may help to uncover the phenomenon of cancer progression. The study for understanding the tumors at the molecular level can be done through microarray analysis which is array of huge amount of biological data on a 2D solid substrate by using high throughput screening method. The main aim of present study is

to understand the similarity and dissimilarity between tumors of epithelial and mesenchymal origin by microarray analysis. Microarray analysis of differentially expressed genes was done in Genespring software and gene ontology study was done in Genomatix software. These analysis lead to the selection of two genes for further study based upon their regulation and association with cancer. Expressions of these genes were confirmed in HeLa cell line which is a cervical cancer cell line taking beta actin (a housekeeping gene) as reference gene through qRT-PCR. This study can lead to possible discovery of genes that can be used biomarker for group of cancers based on their tissue of origin. It might give a clue on cancer metastasis as well. By this, efforts could be made to search for a successful biomarkers and subsequent treatment measure to cure this deadly disease with great efficacy.

REVIEW OF LITERATURE

REVIEW OF LITERATURE

3.1. Tumours

When cells in a part of the body start to grow abnormally and uncontrollably, they are called tumour. They can be benign, premalignant or malignant. The cells of benign tumours are close to normal in appearance and do not have the potential of being dangerous. They grow slowly and do not invade nearby tissues or spread to other parts of the body. Pre-malignant tumors have the potential to become malignant, but the cells do not show the property of uncontrolled proliferation. On the other hand a malignant tumour is cancerous and can spread beyond the original tumor to other parts of the body by invading the normal tissue. One of the critical stages in the progression of malignant tumor is the loss of control of cell cycle regulation (Hartwell and Kastan, 1994). Cancers are named according to the place from where they first started irrespective of where they spread [According to the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)] (fig.1). There are different types of cancer depending on their site of origin and each cancer behaves differently showing variation in their growth rates and response to various treatment measures. The main reason for such variations is the factors which influence the incidence of cancer. These factors include sex, age, diet, genetics and environmental factors (Alberts et al., 2002).

Fig.1. Different kinds of tumors according to ICD-O-3

The malignant tumors are characterized by some features typical to them (Slooten et al., 1985; Nishida et al., 2006). It includes i) Moderate or great changes in the cells or tissues like necrosis, hemorrhage etc. ii) Formation of abnormal structures iii) Moderate or extensive abnormality in the structure of the nucleus and nucleoli iv) Hyperchromatic state of nucleus v) Formation of new blood and lymphatic vessels by the process called angiogenesis or lymphangiogenesis.

Malignant tumors, otherwise known as cancers mainly arise from 4 types of cells which includes epithelial cells, mesenchymal cells, hematopoietic cells (also constitutes cells of immune system) and cells of central and peripheral nervous system. Though there are exceptions like melanomas which don't fit in any of the major classification (Weinberg, 2006).

3.2 Epithelial Tumors

Most of the human tumors are of epithelial origin (Weinberg, 2006). Epithelial cells form the lining of the walls or cavities of the body. In case of skin it acts as the outer covering. Beneath it lays the basal membrane which separates epithelial cells from connective cells. Malignant epithelial tumors are also termed as carcinoma. Carcinoma is responsible for most of the cancer related death all around the world and includes mainly cancers of the mouth, esophagus, stomach, cervix, small and large intestine. It also comprises skin, lungs, liver gall bladder, urinary bladder, and ovarian cancers.

Epithelial tissue arises from three of the primitive cell layer namely ectoderm, endoderm and mesoderm (Chuai et al., 2006). Epithelial cells have shown to have two major biological functions. They are:-

i) To cover and close the cavity and protect the cell population present inside it.

ii) Secretion of substances inside the cavities or ducts around which it form the lining.

Based upon these two function carcinoma are divided into two major categories namely squamous cell carcinoma and adenocarcinoma (Weinberg, 2006).

Squamous cell carcinoma arises from the cells which are mainly associated with formation of protective layer like lining of skin, esophagus etc. Squamous cell carcinoma of skin accounts for majority of skin cancer.

Adenocarcinoma arises from the cells that secrete substances into the duct or cavity or glanular tissue. The main characteristics of adenocarcinoma are its secretory properties and association with gland. The name itself suggests the fact as "adeno" means pertaing to a gland and "carcinoma" means cancer (table1).

TABLE 1: Types of carcinoma and its site of occurrence

Site of occurrence of squamous cell carcinoma	Site of occurrence of adenocarcinoma
Skin	Lungs
Tongue	Stomach
Cervix	Colon
Nasal Cavity	Prostate
Larynx	Pancreas
Oropharynx	Rectum
Lungs	Ovary
Esophagus	Gall bladder
Mouth	Endometrium
Parotid gland	Esophagus
Eyelid	Breast
	Cervix

3.2.1 Stomach Cancer

Stomach cancer or Gastric Carcinoma stands second as the cause of death due to cancer worldwide (Nagini, 2012). The ratio of stomach carcinoma is higher than other common malignancies like breast cancer, prostate cancer etc. (Catalano et al., 2009). Though there is advancement in diagnosis process still majority of patients are not diagnosed at the earlier stage and later suffer from anemia and weight loss, during the advance stage as the body refuses meat based food (Nagini, 2012). It arises from an inflammation and has a long pre-cancer period and then develops into cancer via numerous steps (Carl-McGrath et al., 2007). Stomach adenocarcinoma arises from the stomach mucosa containing neoplastic cells and develops into glanular like tubular structure which frequently ulcerate and is characterized by penetration and thickening of wall of the stomach (Nagini, 2012). It is responsible for 90-95% of all gastric malignancies (Shang and Pena, 2005).

3.2.2 Cervical Cancer

Cervical cancer or cervical carcinoma is a sexually transmitted disease caused by infection of human papillomavirus (HPV) (Okechukwu, 2011). Worldwide study shows that HPV DNA is 99% associated with invading cervical cancer (Parkin, 2006; McLaughlin-Drubin, 2008). It is second most widespread cancer in women (Schiffman et al., 2007). The condition of developing country like India is worse. It is estimated that in India alone the cervical cancer patient were 100,000 in the year 2001(Shanta et al., 2001). Cervical cancer progression involves 4 steps which includes- i) Infection in the cervical transformation zone epithelium ii) Viral perseverance iii) Development of infected epithelium to precancer iv) Progression and invasion (Schiffman et al., 2007).

3.3. Neuroepithelial Tumor

In the year 1993 WHO gave a new and complete classification of neoplasm that affects the central nervous system. The classification was mainly based upon the foundation that tumors results from the anomalous growth of particular cell type. According to this classification neuroepithelial tumors includes: i) Astrocytic tumors (glial tumors) ii) Oligodendroglial tumors iii) Ependymal cell tumors iv) Mixed gliomas v) Neuroepithelial tumors of uncertain origin vi) Tumors of the choroid plexus vii) Neuronal and mixed neuronal-glial tumors viii) Pineal Parenchyma Tumors ix) Tumors with neuroblastic or glioblastic elements (embryonal tumors).

3.3.1. Glioblastoma or Brain Cancer

Glioblastoma is most fatal of all gliomas which are a set of tumors arising from glial cell in the central nervous system and it is a type of astrocytic tumor (Holland, 2000). Glioblastoma is most aggressive as the patient suffering from this disease dies within a year. Due to its complex nature this disease is hard to treat by therapeutic treatments (Holland, 2000). It develops resistance against cell death by switching from mitochondrial oxidative phosphorylation to cytoplasmic glycolysis (Michelakis et al., 2010).

3.4. Mesenchymal Tumor

Mesenchymal tumors are derived from numerous connective tissues present everywhere in the body and share common origin (Weinberg, 2006). Also known as sarcomas, they are

further classified on the basis of the site of the tumor: bone or soft tissue (Lahat et al., 2008). Of these, soft tissue sarcoma (STS) is the collective term used for malignancies arising in muscles, fat, vessels and fibrous tissues (WHO classification). Mesenchymal tumors can develop from tissues like muscle, nerves, fat, fibrous, blood vessels or deep skin tissues. They can be found in any part of the body mostly arms and legs, but can also be formed in the head, neck, trunk area and retroperitonium. These tumors are not common like carcinomas. Soft tissue sarcomas having complex karyotypes do not show characteristic mutation or fusion of genes and includes dedifferentiated and pleomorphic liposarcoma, leiomyosarcoma and myxofibrosarcoma (Barretina et al., 2010).

3.4.1 Liposarcoma

Liposarcoma arises mainly from soft tissue and the retroperitoneum and are classified into 5 types based upon morphological characteristics: well-differentiated, myxoid, dedifferentiated, pleomorphic and round cell (Richard et al., 2010). Among all the soft tissue sarcoma liposarcoma is the most common and accounts for approximately 20% of all mesenchymal malignancies (Dei Tos, 2000). Pleomorphic liposarcoma is an uncommon type of liposarcoma which has recently been characterized properly (Hornick, 2004).

3.4.2 Leiomyosarcoma

Leiomyosarcoma is an uncommon malignant tumor of the smooth muscle cells. It is also found in the uterus and abdomen. There are different types of leiomyosarcoma like pleomorphic leiomyosarcoma, myxoid leiomyosarcoma, differentiated leiomyosarcoma etc. Pleomorphic leiomyosarcoma, due to its morphological feature and biological behavior was recently characterized as a variant and remains controversial (Oda et al., 2001). Myxoid leiomyosarcoma is a rare form of malignant tumor which is mainly found in the uterus but rarely found in other location (Rubin, 2000).

3.4.3 Malignant Fibrous Histiocytoma (MFH)

Malignant fibrous histiocytoma is said to be the common sarcoma of adults but is controversial both in terms of histogenesis and validity (Randall, 2004). The prognosis of MFH is poor (Konishi, 1984). These tumors show no signs of true histiocytic differentiation

and show no definable line of differentiation. World Health Organization classification now classifies MFH as distinct diagnostic category as undifferentiated pleomorphic sarcoma.

3.5. Microarray Gene Expression Analysis

Gene expression profiling is the measurement of expression of many genes at instant and using microarray this measurement can be carried out for previously identified genes. By this a global picture of cellular function can be created. Past few years has experienced a great change in microarrays for analyzing total genome of organisms and developed chips to measure the expression of 40,000 genes (Watson, 1998). This study encourages transparent design and makes the data available publically (Ioannidis, 2009). Microarrays data can be of different types like DNA microarray, RNA microarray, protein microarray, microbe microarray etc. Microarray gene expression data generated are deposited by various research groups in various databases. The various types of microarray database are:

i) Gene Expression Omnibus (GEO) (from National Center for Biotechnology Information)

ii) ArrayExpress (from European Bioinformatics Institute)

iii) Stanford microarray database

iv) caArray at NCI

v) Genevestigator database

By using microarrays one can measure the relative quantities of particular mRNAs in two or more samples of tissues for thousands of gene simultaneously (Kerr et al., 2000). Microarray analysis techniques are used in the interpretation of data generated from experiments carried out on different types of microarrays. This allows the researchers to explore the expression of a large number of genes and also entire genome in a single experiment. With the use of gene expression microarray analysis, the cellular differentiation and oncogenic pathways of malignant tumors could be found and by this biomarkers as well as therapeutic targets for these tumors could also be generated (Baird, 2005).

Different softwares are used to analyze the microarray data. It includes:

i) **Genespring**: It is widely used, commercially available software for microarray analysis.

ii) **R**: It is free software for statistical computing and graphics. It works on a wide variety of platforms like UNIX platforms, Windows and MacOS.

iii) **Bioconductor:** It is an open source software for analyzing genomic data particularly microarray data.

iv) **TM4**: It is a free set of software for microarray analysis and consists of four major applications: Microarray Data Manager (MADAM), Spotfinder, Microarray Data Analysis System (MIDAS), and Multiexperiment Viewer (MeV).

v) **Spotfire:** It is commercially available software for microarray analysis.

3.6. Gene Ontology Analysis

Web based tools like Genomatix, DAVID Bioinformatics Resourses 6.7, National Cancer Institute Nature Pathway Interaction Database and Metacore were used for pathway analysis of the differentially expressed genes. Genomatix is web based software for microarray analysis. Pathway analysis and linkage of genes is carried out with this software (Abu-Asab et al., 2011). The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID) provides comprehensive tool to understand meaning behind huge number of genes. This software provides functional annotation tools to the user and facilitates microarray analysis by analyzing differentially expressed genes (Liu et al., 2013). Matacore allows pathway analysis of OMICS data and helps to have a comprehensive study of genes and related pathway thereby exploring and identifying potential biomarker. It also allows to interpret global changes in the proteome (Chen et al., 2013).

OBJECTIVES

Objective 1:

Microarray expression analysis of tumors of epithelial and mesenchymal origin

Objective 2:

Identification of overlapping sets of differentially expressed genes expressed in tumors of epithelial origin

Objective 3:

Identification of overlapping sets of differentially expressed genes expressed in tumors of mesenchymal origin

Objective 4:

Molecular classification of tumors on the basis of common set of differentially expressed genes irrespective of their organ specific expression

Objective 5:

Identification of enriched molecular pathways

Objective 6:

Validation of expression of two genes that are comparatively highly up or down regulated

MATERIALS AND METHODS

MATERIALS AND METHODS:

4.1. Gene Expression data:

Gene expression is the process through which genetic information is used in the production of a functional gene product which mainly includes protein. But in case of non protein like rRNA or tRNA genes the product produced is functional RNA. The data generated by microarray hybridization experiments which represent the expression of a specific gene are called as gene expression data.

Gene expression data are stored in databases and some of these databases include GEO (Gene Expression Omnibus), GXD (Gene Expression Database), Gene Expression Atlas etc.

For our study the microarray data were retrieved from GEO database which is a public data depository of functional genomics data. The data was retrieved for normal (control) tissues and cancerous (test) tissues from the different GEO series records (GSE). Experimental data for control (Table 2) and test (Table 3) includes:

TABLE 2: EXPERIMENTAL DATA FOR TUMORS OF EPITHELIAL ORIGIN:

Sl. no	CONTROL	TEST
1	Normal human stomach tissue	Primary gastric tumor
2	Normal brain tissue	High grade glioblastoma
3	Normal cervix epithelium	Cervical squamous cell carcinoma

TABLE 3: EXPERIMENTAL DATA FOR TUMOR OF MESENCHYMAL ORIGIN:

Sl. no	CONTROL	TEST
1	Normal human fat tissue	Soft tissue sarcoma (leiomyosarcoma)
2	Normal human fat tissue	Soft tissue sarcoma (MFH:Pleiomorphic)
3	Normal human fat tissue	Soft tissue sarcoma (liposarcoma:Pleiomorphic)

The criterion for selecting these experimental data was their platform. The platform for all the data was same i.e. Affymetrix HG-U133A with GEO Platform accession number GPL96. Based on same platform six tumors were selected with three from epithelial origin and three from mesenchymal origin. The GEO accession number of the data is given in Table 4.

TABLE 4: EXPERIMENTAL DATA AND THEIR GEO ACCESSION NUMBER

Sl.No	Tumor	Control accession number	Test accession number
1	Stomach	GSE2361	GSE15456
2	Brain	GSE13471	GSE8692
3	Cervical	GSE9750	GSE9750
4	Leiomyosarcoma	GSE21122	GSE21122
5	MFH:Pleiomorphic	GSE21122	GSE21122
6	Liposarcoma	GSE21122	GSE21122

4.2 Microarray analysis of gene expression data-

4.2.1 Retrieval of gene expression data-

From the above mentioned GSEs (a series record of GEO that links together a group of related samples and provides a focal point and description of the whole study) individual GEO samples (GSM) were selected. The samples were selected in triplicates; this was done to minimize the error rate. For each of the tumors three control experimental data and three test experimental data were selected. The selected GSM for each tumor is (Table 5):

TABLE 5: SELECTED GSM FOR EACH TUMOR

Sl. No.	Tumor	Control	Test
1	Stomach	GSM44703 GSM44703 GSM44703	GSM387757 GSM387758 GSM387759
2	Brain	GSM339552 GSM339553 GSM339554	GSM215420 GSM215423 GSM215427
3	Cervical	GSM247188 GSM247189 GSM247190	GSM247650 GSM247651 GSM247652
4	Leiomyosarcoma	GSM528425 GSM528426	GSM528323 GSM528324

		GSM528427	GSM528328
5	MFH:Pleiomorphic	GSM528431 GSM528432 GSM528433	GSM528350 GSM528355 GSM528380
6	Liposarcoma	GSM528428 GSM528429 GSM528430	GSM528410 GSM528411 GSM528412

The process of retrieving the data from GEO is:

GEO profile database of NCBI is searched for studies relevant to our interests

Raw data in .CEL format was downloaded from the supplementary files provided with each GSM at the end of the section

After downloading the compressed files, they were unzipped

Files were renamed as control and test

These files were imported into GeneSpring software for microarray data analysis.

4.2.2 Analysis of gene expression data

Genespring GX software (GeneSpring GX 11; Agilent, Santa Clara, CA) was used for analyzing the gene expression data. This software provides powerful and accessible statistical tools for high-speed visualization and analysis of transcriptomics, genomics, proteomics and metabolomics data. It gives the user an interactive computing environment for understanding microarray data within a biological context. Genespring provides an integrative platform for multi-omic data analysis. Key feature of this software includes Gene-level expression analysis on all major microarray platforms, including Agilent, Affymetrix, and Illumina, microRNA analysis and identification of gene targets, correlative analysis on mRNA

expression and miRNA data, real-time PCR data analysis, NCBI Gene Expression Omnibus Importer tool for expression datasets.

Steps followed for gene expression analysis in GeneSpring:

1. The raw data files from GEO database were downloaded as a zip file.
2. Then the files were unzipped, extracted and renamed according to the convenience.
3. GeneSpring was run and a new project with new experiment was created. Experiment name was given then experiment and guided workflow type was selected.
4. Data was uploaded in triplicates with the selected technology as Affymetrix Gene Chip-HG-U113A. Profile plot of Normalized intensity map values is obtained after Normalization of data. Data is normalized to 75th percentile of signal intensity to standardize each chip for cross-array comparison. Normalization is done mainly for eliminating redundancy and ensuring that the data make sense with minimum number of entities.
5. Gene significance/ differentially expressed genes were found through ANOVA using the following parameters: Coefficient of variation <50%, $P < 0.05$, Benjamini-Hochberg false discovery rate as multiple testing correction. Genes were derived using ANOVA and fold change more than 2.0 filtered using the Volcano plot.
6. With the genes with fold change >= 2.0 of tumors of epithelial origin venn diagram is used to obtain the common set of genes between the epithelial tumors. Similarly the common set of genes with fold change >= 2.0 is obtained for mesenchymal tumors. Using these two files sets we obtain the common set of gene between the two tumor type i.e epithelial and mesenchymal and also the exclusive set of genes which are specific for a particular tumor.
7. The common set of genes obtained was exported along with their normalization values, gene symbol and Entrez gene IDs from the GeneSpring software as a tab delimited file for further analysis.
8. Using the obtained common set of genes, clustering analysis was done in GeneSpring using Hierarchical clustering algorithm. The distance metrics used was pearson uncentered and the linkage rule used was average.

4.2.3 Gene Ontology Analysis

The common set of genes was analyzed using GO database (Gene Ontology database), various technologies like Genomatix and web based tools like Web based gene set analysis tool kit and their involvement in various pathways was studied. From the common set of

genes two genes were selected based upon their regulation and association with cancer for further validation by qRT-PCR.

4.3 Experimental Validation

4.3.1 Cell culture

For our study HeLa cell line was taken, these are human cervical cancer cells. HeLa cell line are the first type of human cancer cell which are cultures continuously for experiments. It was first derived from cervical cancer cells in 1951 from a patient named Henrietta Lacks. The medium used for culturing the cell is MEM (Invitrogen; MEM with NEAA (non essential amino acids) and L-Glutamine) with 10% FBS (Fetal bovine serum from HIMEDIA) and 1% antibiotic solution (Penstrep solution from HIMEDIA). The culture flask containing the cell line is kept in the CO_2 incubator with the level of CO_2 maintained at 5%. With the utilization of medium the color of the medium changes from red to orange and then pale yellow because of change in pH of the medium.

The steps for cell culture:

1. The cells were harvested first.

   Cells were grown in suspension i.e. 1 x 107 cells. The number of cells was determined. The appropriate number of cells was pelleted by centrifuging for 5 min at 300 x g in a centrifuge tube. Carefully removed all supernatant by aspiration completely from the cell culture medium.

   To trypsinize and collect cells: The number of cells was determined. The medium was aspirated, and the cells were washed with PBS. Then the PBS was aspirated, and 0.1–0.25% trypsin in PBS was added. After the cells detach from the flask, medium (containing serum to inactivate the trypsin)was added , the cells were transferred to an RNase-free glass or polypropylene centrifuge tube and centrifuged at 300 x g for 5 min. The supernatant was aspirated completely, and proceeded to step 2.
2. The cells was disrupted by adding Buffer RLT:

   For pelleted cells, loosen the cell pellet thoroughly by flicking the tube. 350 µl Buffer RLT was added. Vortexed or pipetted to mix, and ensured that no cell clumps were visible and proceeded to step 3.
3. The lysate was homogenize for 30 s using a rotor–stator homogenizer and proceeded to step 4.

4. 1 volume of 70% ethanol was added to the homogenized lysate, and mixed well by pipetting. Did not centrifuge.
5. 700 µl of each sample was transferred from step 4, including any precipitate to each RNeasy spin column on the vacuum manifold.
6. The vacuum was switched on and was applied until transfer was complete. Then switched off the vacuum and ventilated the vacuum manifold.
7. 700 µl Buffer RW1was added to each RNeasy spin column.
8. The vacuum was switched on and was applied until transfer was complete. Then switched off the vacuum and ventilated the vacuum manifold.
9. 500 µl Buffer RPE was added to each RNeasy spin column.
10. The vacuum was switched on and was applied until transfer was complete. Then switched off the vacuum and ventilated the vacuum manifold.
11. 500 µl Buffer RPE was added to each RNeasy spin column.
12. The vacuum was switched on and was applied until transfer was complete. Then switched off the vacuum and ventilated the vacuum manifold.
13. The RNeasy spin columns was removed from the vacuum manifold, and was placed each in a 2 ml collection tube. The lids were closed gently, and centrifuged at full speed for 1 min.
14. Each RNeasy spin column was placed in a new 1.5 ml collection tube. 30–50 µl RNase free water was added directly to each spin column membrane. The lids were closed gently, and centrifuged for 1 min at 8000 x g (10,000 rpm) to elute the RNA.
15. If the expected RNA yield is >30 µg, then step 15 was repeated using another 30–50 µl RNase free water or using the eluate from step 14 (if high RNA concentration is required). The collection tubes were reused from step 14.

Note: If using the eluate from step 14, the RNA yield will be 15–30% less than that obtained using a second volume of RNase-free water, but the final RNA concentration will be higher.

4.3.2 RNA Isolation

RNA isolation is carried out using QIAGEN kit. The steps involved in the isolation of RNA:-

1. A maximum of 1×10^7 cells were harvested as a cell pellet and the the appropriate volume of Buffer RLT (Lysis buffer) was added.
2. 1 volume of 70% ethanol was added to the lysates and mixed well by pipetting. It should not be centrifuge.

3. 700 µl of the sample, including any precipitation was transfered to an RNeasy Mini spin column placed in a 2ml collection tube (supplied). It was centrifuged for 15s at ≥8000×g (13,000 RPM). Flow –through was discarded.
4. Then 700 µl Buffer RW1 (wash buffer) was added to the RNeasy spin column. It was centrifuged for 15s at ≥8000×g (13,000 rpm). Flow –through was discarded.
5. 500 µl Buffer RPE (wash buffer) was then added to the RNeasy spin column. It was centrifuged for 15s at ≥8000×g (13,000 rpm). Flow –through was discarded.
6. 500 µl Buffer RPE was again added to the RNeasy spin column. It was centrifuged for 15s at ≥8000×g (13,000 rpm). Flow –through was discarded.
7. The RNeasy spin column was then placed in the new 1.5 ml collection tube and 30-50 µl of RNase- free water was added directly to the spin column membrane. It was centrifuged for 15s at ≥8000×g (13,000 rpm) to elute the RNA**.**
8. Using eppendorf nanodrop the purity and quantity of RNA yield is checked by taking only 1-2µl of the sample.

Precaution:

1. A clean and sterilized environment was maintained while working with the RNA as there is high chance of contamination by RNase which may degrade the RNA and lead to reduced RNA yield.
2. When working with chemicals, always wear a suitable lab coat and disposable gloves.
3. Contact of chemicals with acids should be prevented as the combination liberates very toxic gas.
4. In case of contact with eyes, rinse immediately with plenty of water and seek medical advice.
5. The chemicals are harmful if inhaled or if comes in contact with skin and if swallowed.

4.3.3 cDNA synthesis

cDNA synthesis was carried out using SuperScript First-Strand Synthesis System for RT-PCR by Invitrogen using oligo dT primers.

The steps in cDNA synthesis:

1. Each component of the kit was mixed and briefly centrifuge before use.
2. For each reaction, the following components were combined in a sterile 0.2 or 0.5ml tube.

Components	**Amount**
RNA (2µg)	n µl
10 mM dNTP mix	1 µl
Primer (0.5µg/µl oligo $(dT)_{12-18}$)	1µl
DEPC treated water	To 10µl

3. The RNA/primer mixture was incubated at 65·c for 5 minutes and then placed on ice for at least 1 minute.
4. In a separation tube, the following 2X reaction mix was prepared, by adding each component in the indicated order.

Components	**1RXn**	**10 RXnS**
10X RT buffer	2 µl	20 µl
25mM $Mgcl_2$	4 µl	40 µl
0.1M DTT	2 µl	20 µl
RNase out ™ (400/ µl)	1 µl	10µl

5. 9µl of the 2X reaction mixture was added to each RNA/primer mixture from step 3 and mixed gently and collect by briefly centrifugation.
6. It was incubated at 42·c for 2 minutes.
7. 1µl of super script™ II RT was then added to each tube
8. Incubated at 42·c for 50 minutes

9. Then, the reaction is terminated at 70·c for 15 minutes and chilled on ice

10. The reaction was collected by brief centrifugation and 1µl of RNase H was added to each tube and incubated for 20 minutes at 37·c.

11. Then, the reaction is stored at -20·c or used for PCR immediately.

4.3.4 qRT-PCR

Real time PCR is a method that allows exponential amplification of DNA sequences and simultaneously quantifies it. This system is based on the detection and quantitation of a fluorescent probes. Probes which are used in qRT-PCR are taqman probes, molecular beacon , SYBR® Green, displacing probes, light up probes etc. For present study we used SYBR® Green probe which is a frequently used fluorescent DNA binding probe and relies on the sequence specific detection dye. The genes which are taken for the experiment are of two types test genes and reference genes. Reference genes are used to compare the expression of test genes i.e. how much fold they have increased or decreased with respect to normal expression. Reference genes should have following feature: the standard gene should have the same copy number in all cells, it should be expressed in all cells, a medium copy number is advantageous since the correction should be more accurate. For this experiment the test genes were ABCA8 and SMC4 and reference gene used is beta-actin. Beta actin is a house keeping gene also called as constitutive genes which are required for the maintenance of fundamental cellular function, and are expressed in all cells of an organism under normal and patho-physiological conditions. Cycle temperature and time is mentioned in Table.6. and fig.2.

Table 6: Cycle Temperature and time for real time PCR

STAGE	TEMPERATURE (°C)	TIME	CYCLE
Stage 1	95	20 sec	1
Stage 2	95	15 sec	40
	55	15 sec	
	68	20 sec	
Stage 3	95	15 sec	1
	60	15sec	
	95	15 sec	

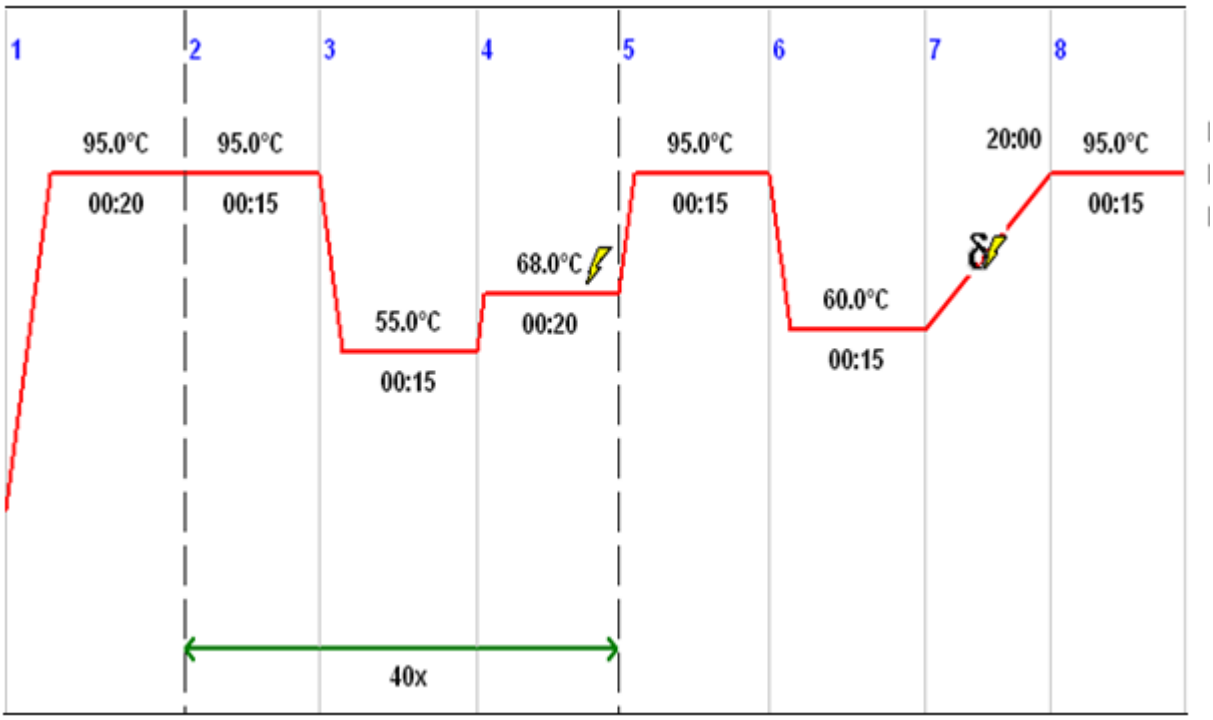


Fig.2. Cycle temperature and time of qRT-PCR

The The total reaction volume could either be 5µl or 10µl. We prepared total reaction volume of 10µl.

Calculation:

We took 3 genes, therefore,

3 genes x 3 replicates = 9 x 10 = 90µl

=100µl (in case of pipetting error do addional 10 µl is taken)

i) **SYBR ® Green master mix:-**

The stock solution is 2X concentration and working solution of 1 X concentration is prepared.

2X x (? µl) = 1X x 100µl

⇨ (? µl) = 1X x 100µl / 2X

= 50µl

SYBR ® Green master mix= 50µl

ii) **cDNA:**

cDNA = cDNA stock : distilled water

1 : 20

cDNA=3µl

Distilled water = 57µl

Total= 3+57 µl=60µl

For each reaction we require 4µl of cDNA

Therefore, for 10 reaction= 4x10= 40µl

From the above made 60 µl we took 40µl

iii) **Primer:-** The stock solution contains 10µM, we require 500nM for each reaction

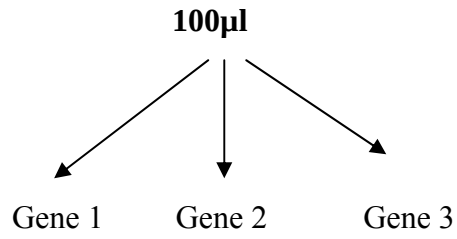
10 µM x (? µl) = 500x1000µMx100

⇨ (? µl) = 500x1000µMx100 / 10 µM

⇨ = 5µl

For forward and reverse primer 5x2=10µl

SYBR ® Green master mix + cDNA + Primer= (50 +40+10) µl= 100 µl



For each gene 3 replicates is taken

The primer sequences were obtained from Primer Bank Harward and ordered from Sigma Genosys and sequence for each gene is**:**

Reference gene:	**Amlicon Size**
Beta actin:	**250**

Forward primer- CATGTACGTTGCTATCCAGGC

Reverse primer- CTCCTTAATGTCACGCACGAT

Test Gene:

a) ABCA8: **176**

Forward primer- TTCATGTTGGCATTTGACACTTG

Reverse primer- GGATCGGCATCCATTTCATCT

b) SMC4: **87**

Forward primer- CGCCTCCAGCAATGACCAAT

Reverse primer- CCCCAGCATAGGATTTGAAGTT

RESULTS AND DISCUSSIONS

RESULTS AND DISCUSSIONS:

5.1 Microarray analysis

From microarray analysis using Genespring, we got differentially set of genes for each of the cancer types considered in our study with fold change (FC) >=2.0. Thereafter, we obtained 344 differentially expressed genes (up/down) which are common among the tumors of epithelial origin.

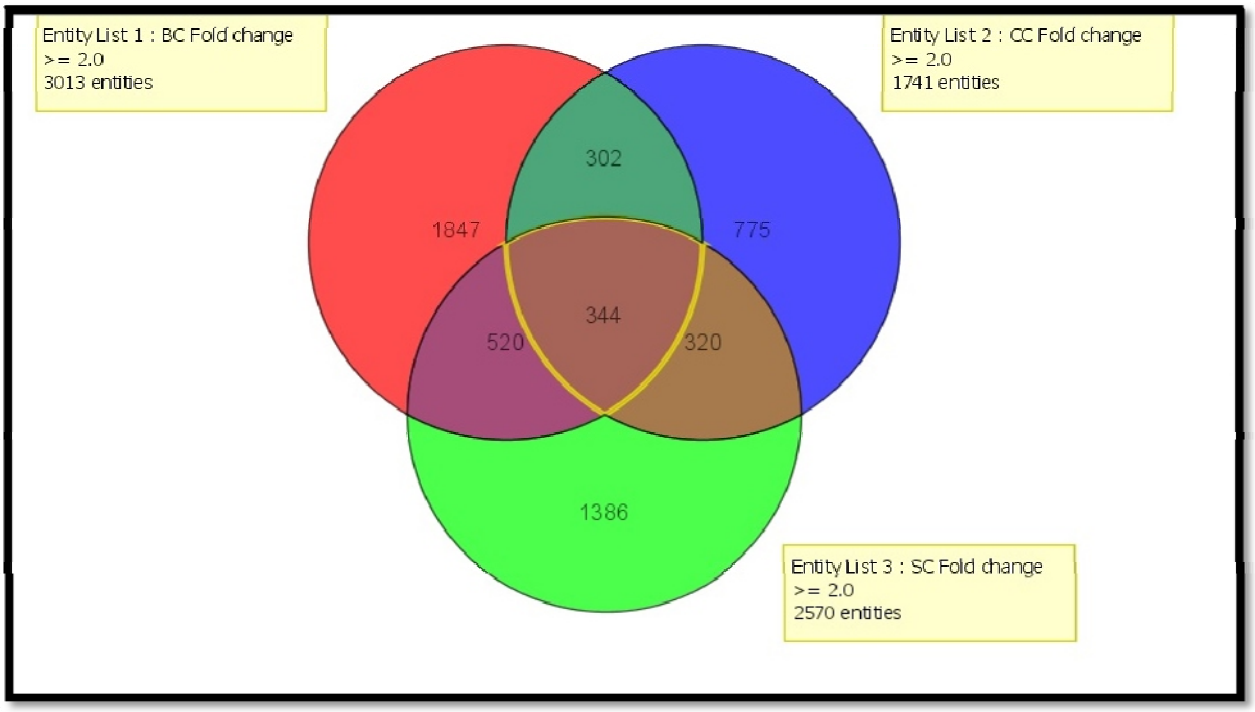


Fig.3. Venn diagram showing common set of differentially expressed genes among epithelial tumors

From this analysis, we not only got the common set of genes between tumors of epithelial origin, but also got the exclusive genes that are expressed in brain, cervical and stomach cancer.

Similarly, we got 672 genes which are differentially expressed in all tumors of mesenchymal origin.

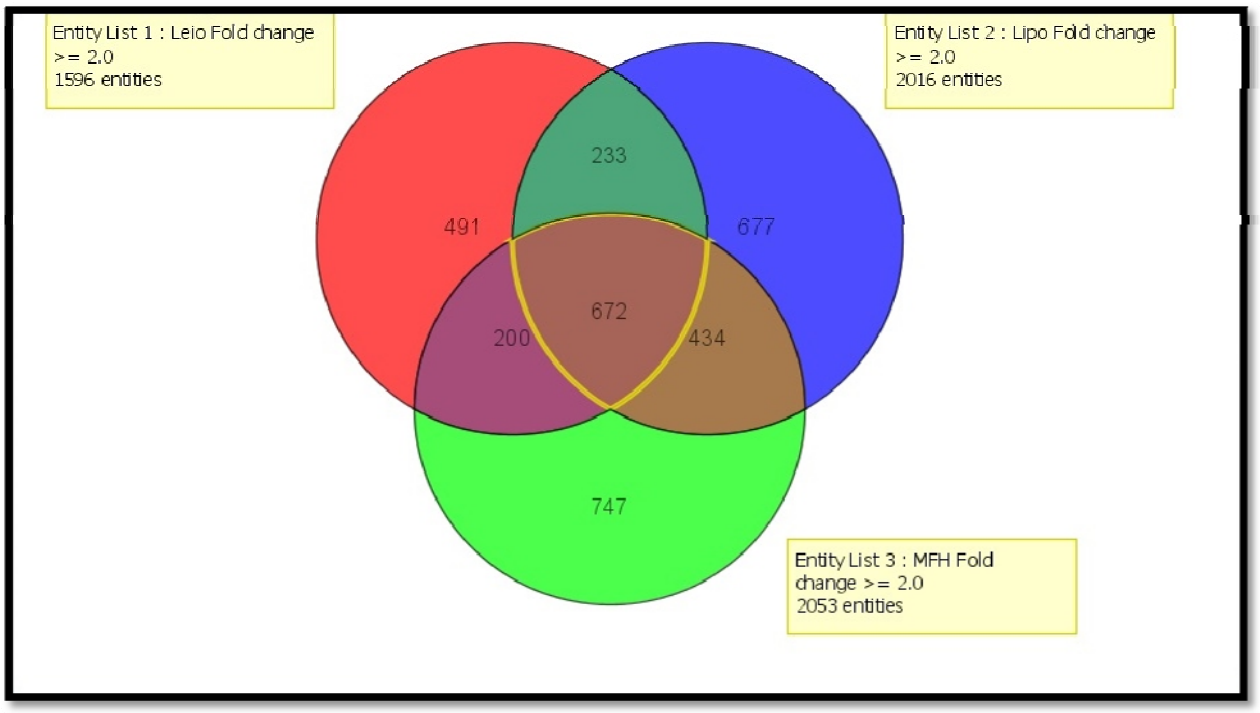


Fig.4. Venn diagram showing common set of differentially expressed genes among mesenchymal tumors

From the common set of differentially expressed genes between tumors of epithetlial and mesenchymal origin, we got the genes which are common to both epithelial and mesenchymal tumors.

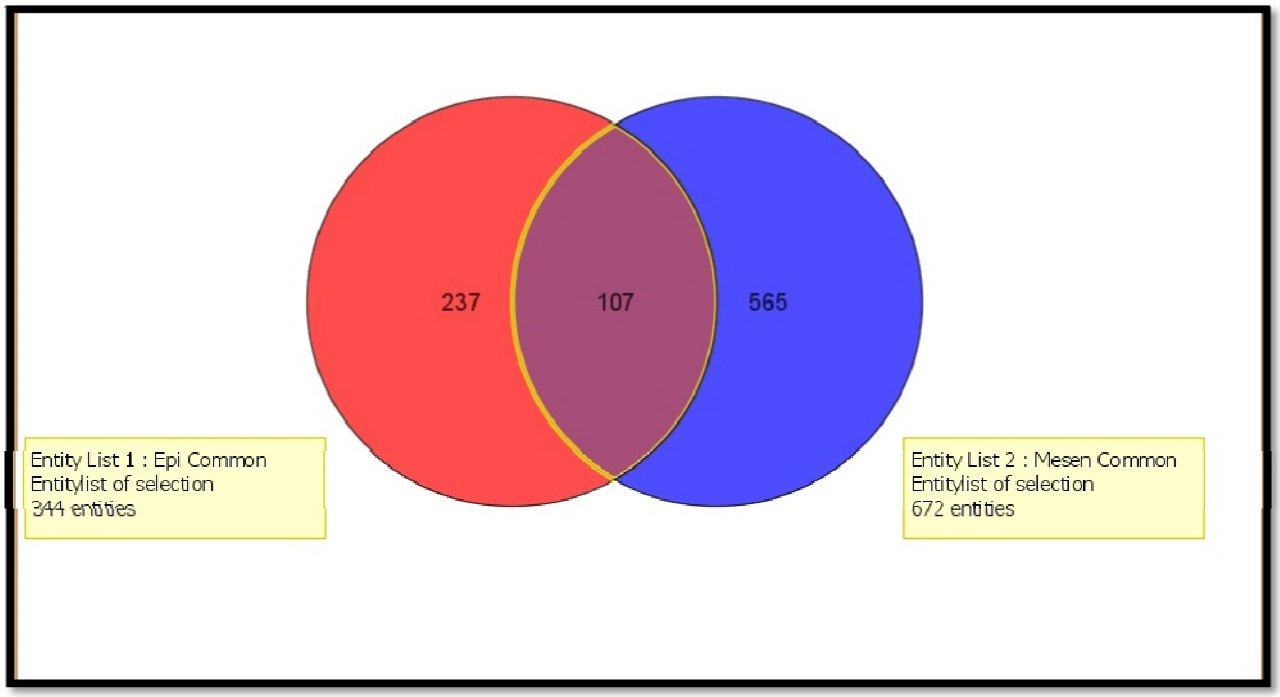


Fig.5. Common set of genes between epithelial and mesenchymal tumors

Total of 107 genes were obtained which upon further screening, i.e. removal of redundant data or similar probe set id, yielded 89 genes which are common to both tumors of epithelial origin and mesenchymal origin.

Considering the 89 common genes and their normalization intensity values, we did the hierarchical clustering of these six tumors and obtained the following dendrogram. The dendrogram shows that cervical and stomach cancer clustered together, but brain cancer clustered separately it may be due to the neuroepithelial origin of brain cancer. Even though brain cancer is grouped differently than other cancers, it is still closer to rest two cancers of epithelial origin. Tumors of mesenchymal origin clustered together.

The classification mainly formed two clusters one of epithelial tumors and other of mesenchymal tumors.

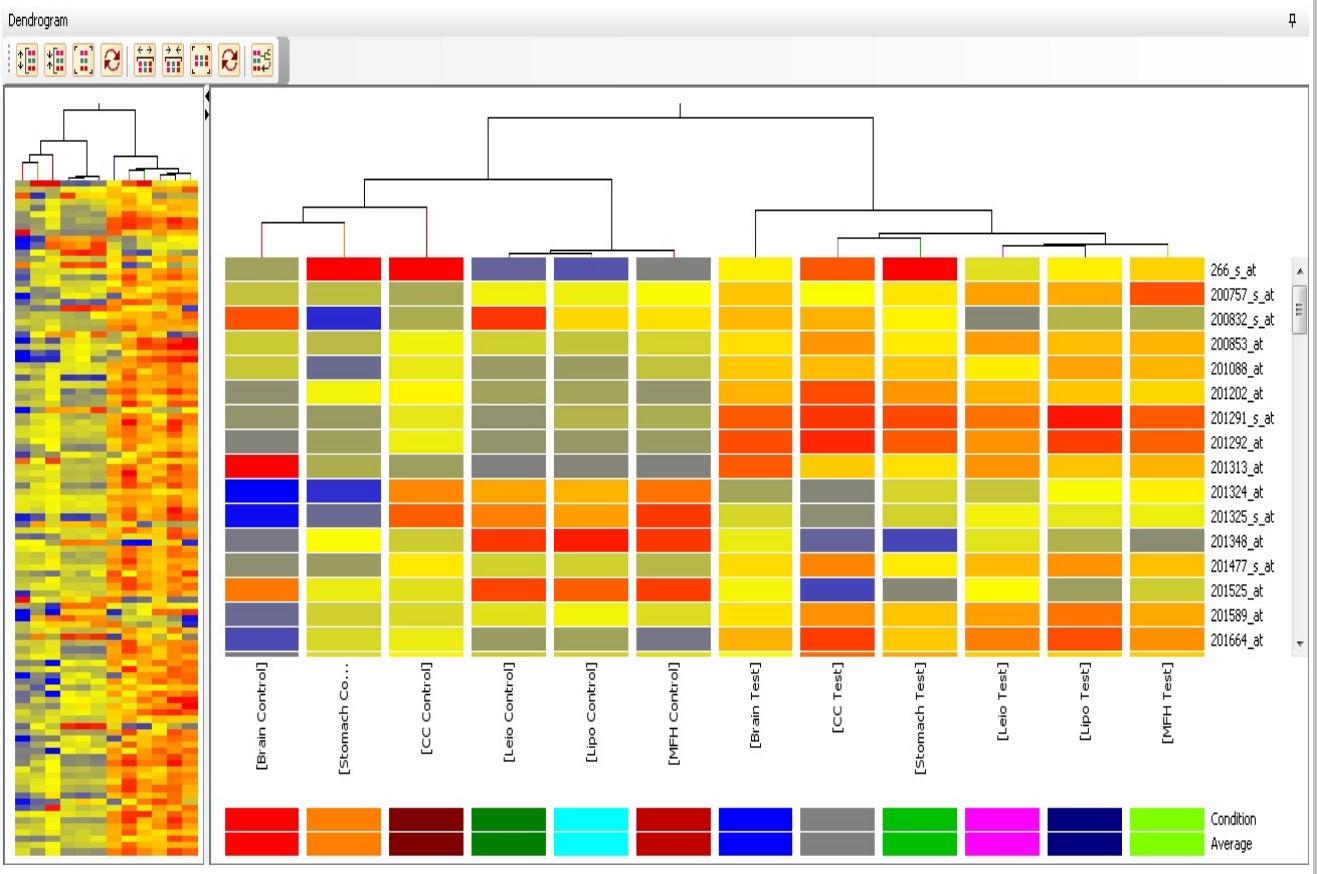


Fig.6. Clustering of common differentially expressed genes

5.2. Gene Ontology Analysis

The 89 genes were analyzed using genomatrix to see if they belong to same or common pathway.

From the analysis we got two enriched pathway: PLK1 pathway and FOXM1 transcription network

a) Polo Like Kinase 1 (PLK1) pathway:

- PLK1 was found to have a critical role in many cancers (both epithelial and mesenchymal origin) because of its role in mitotic progression

15 genes from the common set of genes were present in PLK1 pathway.

Analysis of 15 genes that was present in enriched PLK1 pathway showed that the regulation of each of the 15 genes were similar in both epithelial and mesenchymal tumor.

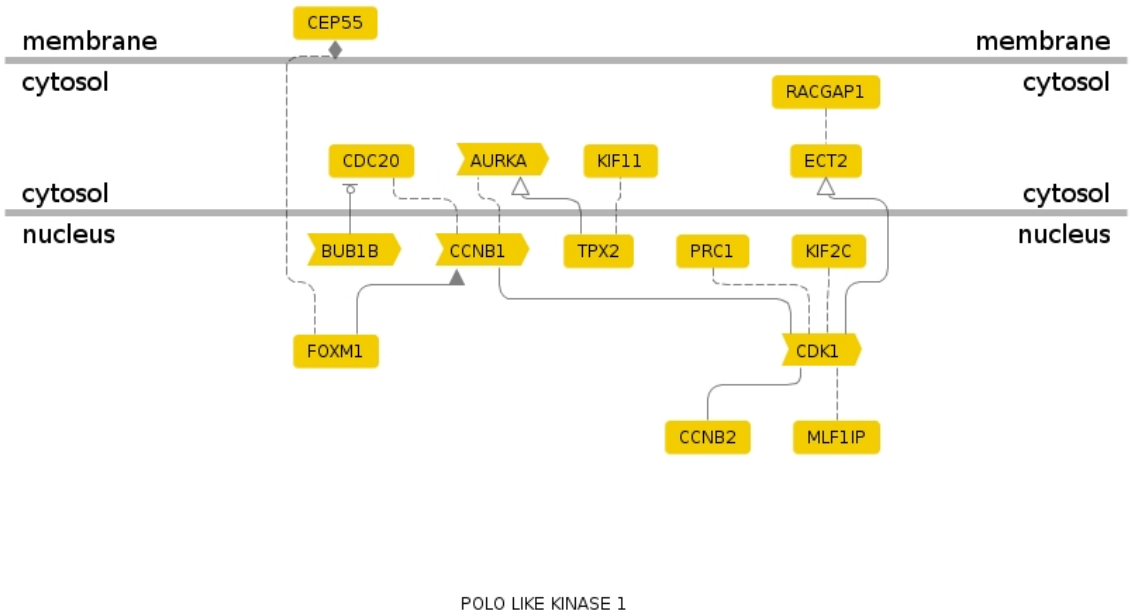
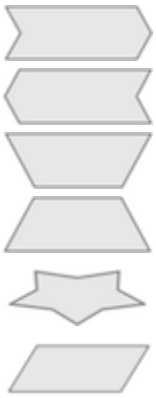


Fig.7. PLK1 pathway having 15 genes from common set of 89 genes

By default a **gene product** is drawn as rounded rectangle



kinase

phosphatase

receptor

transporter

co-factor

RNA

2 genes are associated by co-citation.

2 genes are associated by expert-curation.

Gene A activates Gene B.

Gene A inhibits Gene B.

Gene A modulates Gene B.

Gene A alters the state of Gene B.

If gene A has a known TF binding site matrix and gene B has a corresponding **binding site** in one of its promoters the arrow is filled black. For interactions that involve a complex, this arrow type is never used. To look for promoter bindings in this case, double-click on the edge and select the interaction of interest.

There is no promoter binding noted

b) FOXM1 transcription network

- FOXM1 , a transcription factor was also found to be regulated by pathways involved in cancers

7 genes from the common set of genes were present in the FOXM1 transcription network(fig.6).

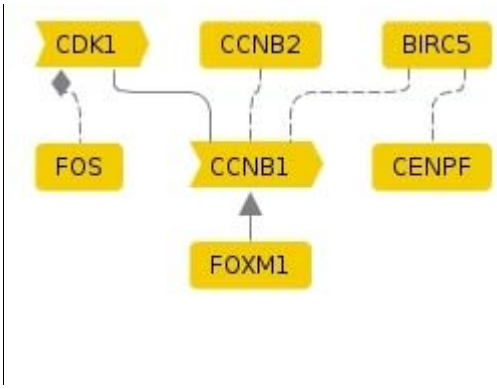


Fig.8. FOXM1 transcription factor network having 7 genes from common set of 89 genes

From the pool of 89 the common genes that are differentially expressed in both tumors of epithelial & mesenchymal origin, we selected only two genes (ABCA8 and SMC4) for further study and experimental validation in cancer cell lines.

The official full name of ABCA8 is ATP-binding cassette, sub-family A (ABC1), member 8. It is a membrane-associated gene which encodes for a member of the superfamily of ATP-binding cassette (ABC) transporters. ABCA8 is a member of ABC1 subfamily and found exclusively in multicellular eukaryotes. The function of this gene is yet to be discovered. Direct association of ABCA8 with cancer is not seen but it is seen to be downregulated in drug resistant ovarian cancer cell line (Januchowski et al, 2013), in tongue squamous cell carcinoma (Ye et al., 2008).

The official full name of SMC4 is structural maintenance of chromosomes 4. It is a member of structural maintenance of chromosomes, or SMC, family. They are responsible for mitotic chromosome condensation in frogs and for DNA repair in mammals. Studies suggest that SMC4 is associated with tumorigenesis. Overexpression of the SMC 4 protein is associated with tumor de-differentiation, advanced stage and vascular invasion of primary liver cancer (Zhou et al., 2012).

Both the genes have function in ATP binding. They are associated with respiratory diseases like pneumonia.

From genespring analysis we got that ABCA8 was upregulated and SMC4 was downregulated.

Interaction network of all the 89 common genes with two of our gene of interest is described below (fig.7):

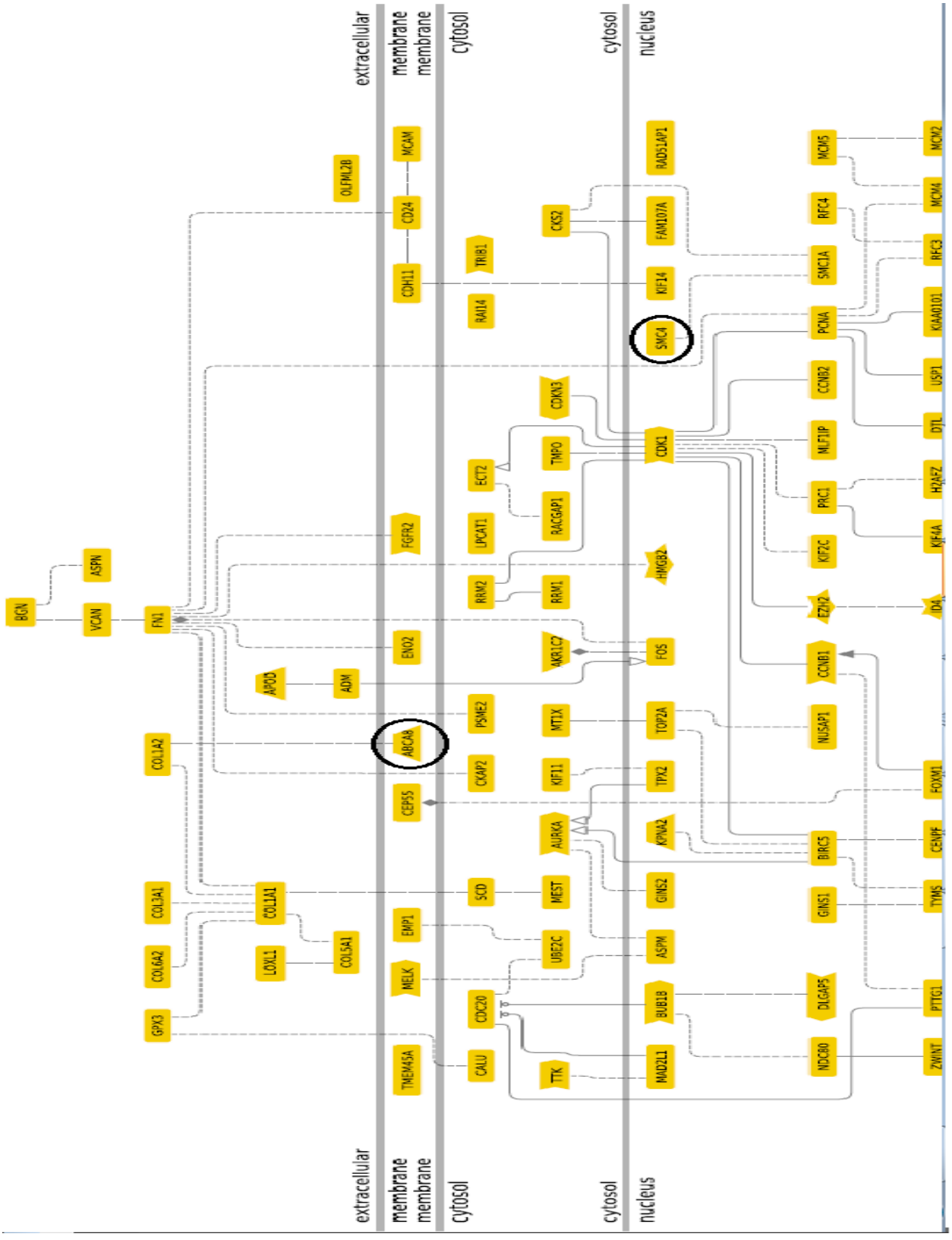


Fig.9. Interaction map of 89 genes with our 2 genes of interest

5.3 Experimental validation

5.3.1. RNA Isolation

The yield after RNA isolation was:

RNA conc. – 498.3µl

2.00-260/280

2.00-260/230

260/280

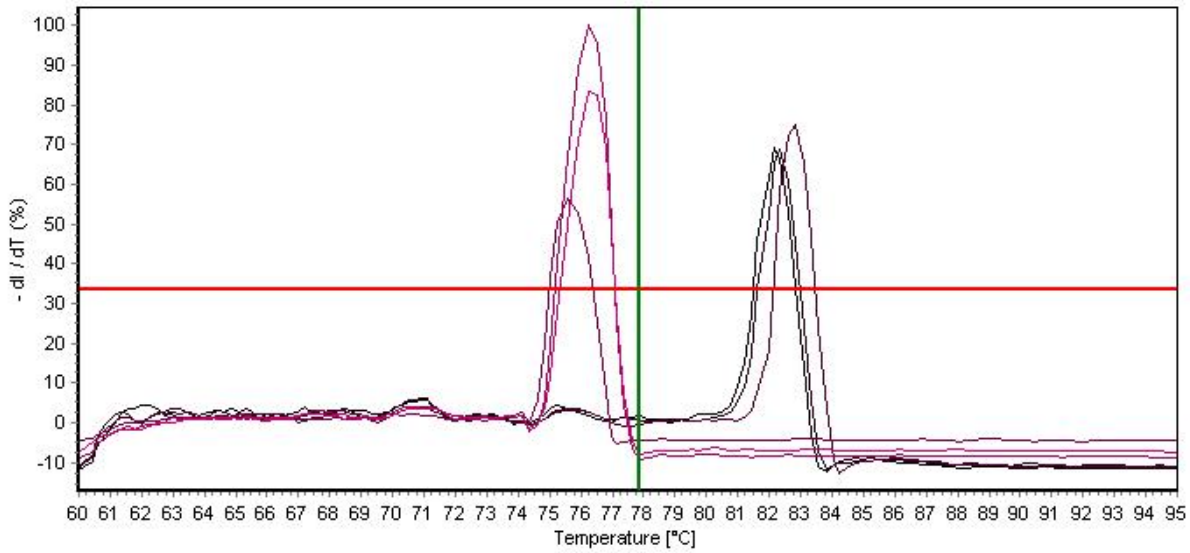
The ratio of absorbance at 260 nm and 280 nm is used to measure the purity of DNA and RNA. A ratio of ~1.8 is usually accepted as "pure" for DNA; a ratio of ~2.0 is usually accepted as "pure" for RNA. If the ratio is significantly lower in each case, it may show the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm.

260/230

This ratio is used as a secondary measure of nucleic acid purity. The 260/230 values for "pure" nucleic acid are frequently higher than the respective 260/280 values. 260/230 values are normally in the range of 2.0-2.2. If the ratio is significantly lower than expected, it may indicate the presence of contaminants which absorb at 230 nm. EDTA, carbohydrates and phenol all have absorbance near 230 nm.

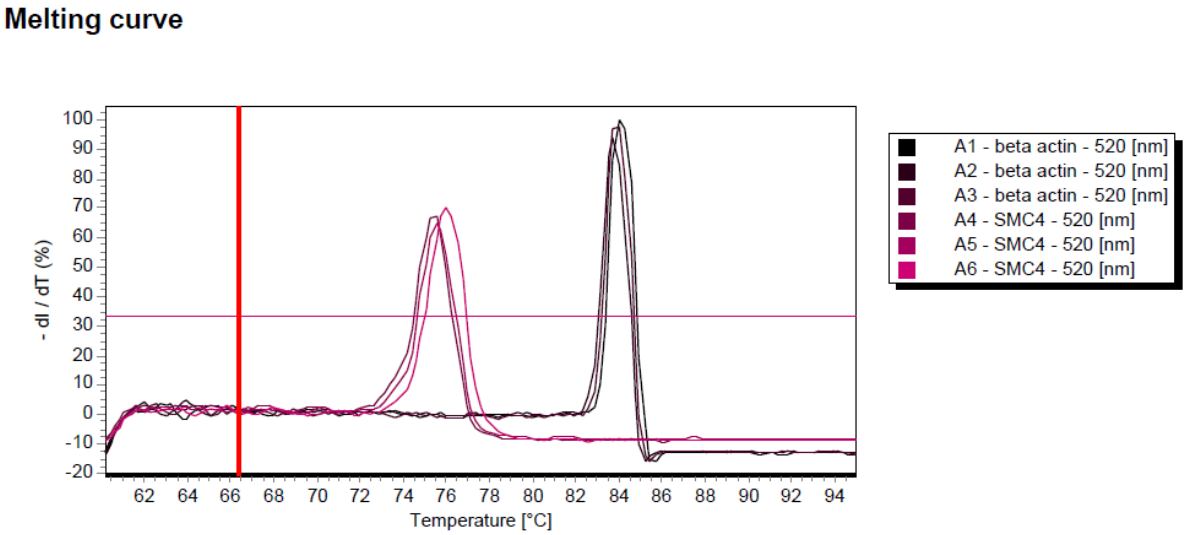
5.3.2. qRT PCR

Melting curve analysis is an assessment of the dissociation of ds DNA during the process of heating. As the temp increases the double strand of the DNA starts to separate leading to a rise in the absorbance. The temp at which 50% of DNA is denatured is known as the melting point. The graphical curve obtained from the real time PCR because of the 50% denaturation of DNA is known as melting temperature curve. The melting temperature curve of ABCA8 and SMC4 is shown below:



Threshold: 33%

Fig.10. Melting temperature curve of ABCA8 and beta-actin



Threshold 33%

Fig.11. Melting temperature curve of SMC4 and beta-actin

Relative quantification is the measure of increase or decrease in fold change with respect to the reference or control gene.

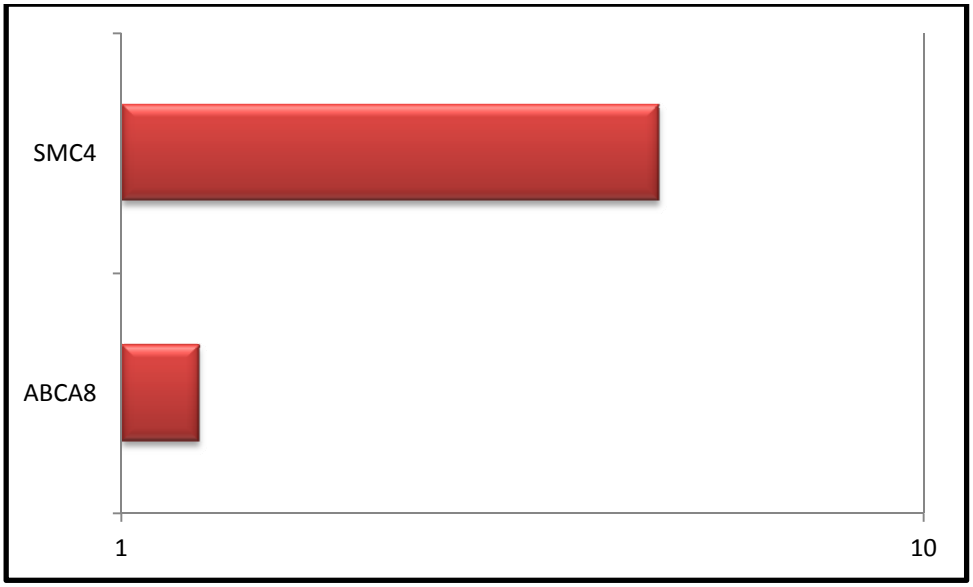


Fig.12. Relative expression of SMC4 and ABCA8 with respect to control

qRT- PCR confirms the presence of SMC4 and ABCA8. Both the genes were found to be upregulated in qRT-PCR results. SMC4 was found to be highly upregulated with respect to control gene beta-actin.

CONCLUSION:

Our microarray analyses of samples of cancers of different tissue origin lead to sets of differentially expressed genes. We found 344 genes are differentially expressed in tumors of epithelial origin, whereas 672 genes in tumors of mesenchymal origin. Our subsequent analysis lead to 89 genes that was common to both the tumor types. Based on the normalization intensity values of these 89 genes, we were able to cluster the two malignant tumors. Cluster analysis of cancers of epithelial and mesenchymal origin using expression values of these common genes shows their classification according to their tissue of origin. This is in accordance with the International classification of Diseases of oncology which classifies cancers according to morphology and topography. Brain cancer, which originates from epithelial cells forms a different group in our clustering result, which might be because of its neuro-epithelial nature. These two tumors share similarity as well as dissimilarity among them which is evident from our study through the clustering analysis.

Gene ontology study of 89 genes through genomatix software gave us two enriched molecular pathway, PLK1 signaling pathway and FOXM1 transcription signaling pathway which contain 15 and 7 genes respectively. These two pathways are reported to play critical role in many malignant tumors. Furthermore, we selected two genes (ABCA8 and SMC4) for experimental validation based on their fold change & hypothesis that these two genes might have critical role in cancer. ABCA8 is ATP-binding cassette sub-family A (ABC1), member 8. The function of this gene is yet to be discovered and SMC4 is structural maintenance of chromosomes 4. They are responsible for mitotic chromosome condensation in frogs and for DNA repair in mammals. Validation of expression of the two genes- ABCA8 and SMC4 by qRT PCR confirmed their expression in the HeLa cell line (cervical cancer cell lines). Further functional annotation of these genes as well validation of these genes in other cancer cell lines might provide deeper insight into their role in cancer biology.

These 89 differentially expressed common genes in all the six malignant tumors and normal cells or subset of it can be used as possible biomarkers for cancers. By using standardized gene expression profile, one can predict cancer types and/or their tissue of origin even though it has been metastasized to other location. These gene profiles can be used as molecular signature to predict a medical condition of a cancer without prior knowledge of their tissue of origin or disease-specific symptoms.

REFERENCES:

1. Weinberg R.A. The Biology of cancer. Garland Science 2006: p. 26-32
2. WHO classification of soft tissue tumours
3. Vijayalakshmi D, Fathima S, Ramakrishnan K, Devi M. Malignant fibrous histiocytoma of the gingival. BMJ case reports 2012
4. Kocak Uzel E, Figen M, Bek TT, Inanc K, Onder S, Kizilkaya HO. Malignant fibrous histiocytoma of the breast in young male patient: a case report and a review of the literature. Case reports in oncological medicine 2013; 2013:524305
5. Wilhite SE, Barrett T. Strategies to explore functional genomics data sets in NCBI's GEO database. Methods in molecular biology 2012; 802:41-53
6. Hartwell LH, Kastan MB. Cell cycle control and cancer. Science 1994; Vol. 266 no. 5192 pp. 1821-1828
7. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. The Preventable Causes of Cancer. Available from: http://www.ncbi.nlm.nih.gov/books/NBK26897/
8. Nishida N, Yano H, Nishida T, Kamura T, and Kojiro M. Angiogenesis in Cancer. Vasc Health Risk Manag. 2006 September; 2(3): 213–219
9. Patricia S, Das S, Rajesh B, Rajesh I, Selvamani B and Subhashini J. Rare Cause of Stricture Esophagus—Sarcoma: A Case Report and Review of the literature. Case Reports in Gastrointestinal Medicine 2011 Volume 2011, Article ID 192423
10. Nagini S. Carcinoma of the stomach: A review of epidemiology,pathogenesis, molecular genetics and chemoprevention World J Gastrointest Oncol 2012 July 15; 4(7): 156-169
11. Catalano V, Labianca R, Beretta GD, Gatta G, de Braud F,Van Cutsem E. Gastric cancer. Crit Rev Oncol Hematol 2009; 71: 127-164
12. Carl-McGrath S, Ebert M, Röcken C. Gastric adenocarcinoma: epidemiology, pathology and pathogenesis Cancer Therapy 2007; Vol 5, 877-894
13. Shang J, Peña AS. Multidisciplinary approach to understand the pathogenesis of gastric cancer. World J Gastroenterol 2005;11(27):4131-4139
14. Okechukwu A.I. Molecular Pathogenesis of Cervical Cancer. Cancer Biology and Therapy, 2011; 11:3, 295-306
15. Parkin D. The global health burden of infection-associated cancers in the year 2002. Int J Cancer 2006;118:3030-44

16. McLaughlin-Drubin M, Munger K, Viruses associated with human cancer. Biochim Biophys Acta 2008; 1782:127-50
17. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007 Sep 8;370(9590):890-907.
18. Shanta V, Krishnamurthi S, Gajalakshmi CK, Swaminathan R, Ravichandran K. Epidemiology of cancer of the cervix: global and national perspective. J Indian Med Assoc. 2001 ;98(2):49-52.
19. Holland E.C. Glioblastoma multiforme: The terminator. Proc Natl Acad Sci U S A. 2000 June 6; 97(12): 6242–6244
20. Michelakis ED, Sutendra G, Dromparis P, Webster L, Haromy A, Niven E, Maguire C, Gammer TL, Mackey JR, Fulton D, Abdulkarim B, McMurtry MS, Petruk KC. Metabolic modulation of glioblastoma with dichloroacetate. Sci Transl Med. 2010 May 12;2(31):31ra34
21. Barretina J. Taylor BS. Banerji S et al. Subtype-specific genomic alterations define new targets for soft tissue sarcoma therapy. Nat Genet. 2010 August ; 42(8): 715–721.
22. Richard SY, Joseph B, Kathleen S. B, Hameed M. Dedifferentiated Liposarcoma of Thigh With Chondrosarcomatous Dedifferentiated Component. Am J Orthop. 2010;39(11):E114-E118
23. Dei Tos AP. Liposarcoma: new entities and evolving concepts. Ann Diagn Pathol. 2000 Aug;4(4):252-66
24. Hornick JL, Bosenberg MW, Mentzel T, McMenamin ME, Oliveira AM, Fletcher CD. Pleomorphic liposarcoma: clinicopathologic analysis of 57 cases. Am J Surg Pathol. 2004 Oct;28(10):1257-67.
25. Oda Y, Miyajima K, Kawaguchi K, Tamiya S, Oshiro Y, Hachitanda Y, Oya M, Iwamoto Y, Tsuneyoshi M. Pleomorphic leiomyosarcoma: clinicopathologic and immunohistochemical study with special emphasis on its distinction from ordinary leiomyosarcoma and malignant fibrous histiocytoma. Am J Surg Pathol. 2001 Aug;25(8):1030-8
26. Rubin BP. Myxoid leiomyosarcoma of soft tissue, an underrecognized variant. Am J Surg Pathol. 2000 Jul;24(7):927-36
27. Randall RL, Albritton KH, Ferney BJ, Layfield L. Malignant fibrous histiocytoma of soft tissue: an abandoned diagnosis. Am J Orthop (Belle Mead NJ). 2004 Dec;33(12):602-8.
28. Konishi Y. Mii Y. Maruyama H and Masuhara K. Malignant fibrous histiocytoma. Am J Pathol. 1984 June; 115(3): 469–472.

29. Watson A, Mazumder A, Stewart M, Balasubramanian S. Technology for microarray analysis of gene expression. Curr Opin Biotechnol. 1998 Dec;9(6):609-14
30. Baird K, Davis S, Antonescu CR, Harper UL, Walker RL, Chen Y, Glatfelter AA, Duray PH, Meltzer PS. Gene expression profiling of human sarcomas: insights into sarcoma biology. Cancer Res. 2005 Oct 15;65(20):9226-35.
31. Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. J Comput Biol. 2000;7(6):819-37. PubMed PMID: 11382364
32. Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E., and Thun, M.J. (2003). Cancer statistics, 2003. CA Cancer J Clin *53*, 5-26.
33. Chuai M, Hughes D, Weijer CJ. Collective epithelial and mesenchymal cell migration during gastrulation. Curr Genomics. 2012 Jun;13(4):267-77. doi:10.2174/138920212800793357.
34. Zhou B, Yuan T, Liu M, Liu H, Xie J, Shen Y, Chen P. Overexpression of the structural maintenance of chromosome 4 protein is associated with tumor de-differentiation, advanced stage and vascular invasion of primary liver cancer.Oncol Rep. 2012 Oct;28(4):1263-8. doi: 10.3892/or.2012.1929. Epub 2012 Jul 24.
35. Ye H, Yu T, Temam S, Ziober BL, Wang J, Schwartz JL, Mao L, Wong DT, Zhou X.Transcriptomic dissection of tongue squamous cell carcinoma. BMC Genomics. 2008 Feb 6;9:69. doi: 10.1186/1471-2164-9-69.
36. Januchowski R, Zawierucha P, Andrzejewska M, Ruciński M, Zabel M. Microarray-based detection and expression analysis of ABC and SLC transporters in drug-resistant ovarian cancer cell lines. Biomed Pharmacother. 2013 Apr;67(3):240-5. doi: 10.1016/j.biopha.2012.11.011. Epub 2013 Jan 1.
37. Abu-Asab M, Zhang M, Amini D, Abu-Asab N, Amri H. Endometriosis gene expression heterogeneity and biosignature: a phylogenetic analysis. Obstet Gynecol Int. 2011;2011:719059. doi: 10.1155/2011/719059. Epub 2011 Dec 13
38. Liu M, Hou X, Zhang P, Hao Y, Yang Y, Wu X, Zhu D, Guan Y. Microarray gene expression profiling analysis combined with bioinformatics in multiple sclerosis.Mol Biol Rep. 2013 May;40(5):3731-7. doi: 10.1007/s11033-012-2449-3.
39. Chen CL, Lin TS, Tsai CH, Wu CC, Chung T, Chien KY, Wu M, Chang YS, Yu JS, Chen YT. Identification of Potential Bladder Cancer Markers in Urine by Abundant-Protein Depletion Coupled with Quantitative Proteomics. J Proteomics. 2013 Apr 27.